Math (P)refresher Lecture 8: Unconstrained Optimization

September 2006

Today's Topics*: • Quadratic Forms • Definiteness of Quadratic Forms • Maxima and Minima in $\mathbf{R}^n$ • First Order Conditions • Second Order Conditions • Global Maxima and Minima

1 Quadratic Forms

- Quadratic forms important because
  1. Approximates local curvature around a point — e.g., used to identify max vs min vs saddle point.
  2. Simple, so easy to deal with.
  3. Have a matrix representation.
- **Quadratic Form**: A polynomial where each term is a monomial of degree 2:

$$Q(x_1, \cdots, x_n) = \sum_{i \leq j} a_{ij} x_i x_j$$

  which can be written in matrix terms

$$Q(\mathbf{x}) = \begin{pmatrix} x1 & x2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \cdots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{1n} & \frac{1}{2}a_{2n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

  or

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

- Examples:
  1. Quadratic on $\mathbf{R}^2$:

$$\begin{aligned} Q(x_1, x_2) &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 \end{aligned}$$

  2. Quadratic on $\mathbf{R}^3$:

$$\begin{aligned} Q(x_1, x_2, x_3) &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \frac{1}{2}a_{13} \\ \frac{1}{2}a_{12} & a_{22} & \frac{1}{2}a_{23} \\ \frac{1}{2}a_{13} & \frac{1}{2}a_{23} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 \end{aligned}$$

*Much of the material and examples for this lecture are taken from Simon & Blume (1994) *Mathematics for Economists* and Ecker & Kupferschmid (1988) *Introduction to Operations Research*.

2 Definiteness of Quadratic Forms

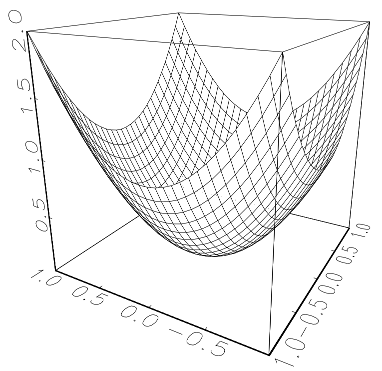
- Definiteness helps identify the curvature of $Q(\mathbf{x})$ at $\mathbf{x}$.
- **Definiteness**: By definition, $Q(\mathbf{x}) = 0$ at $\mathbf{x} = 0$. The definiteness of the matrix $\mathbf{A}$ is determined by whether the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ is greater than zero, less than zero, or sometimes both over all $\mathbf{x} \neq 0$.

1. Positive Definite	$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0, \quad \forall \mathbf{x} \neq 0$	Min
2. Positive Semidefinite	$\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0, \quad \forall \mathbf{x} \neq 0$	
3. Negative Definite	$\mathbf{x}^T\mathbf{A}\mathbf{x} < 0, \quad \forall \mathbf{x} \neq 0$	Max
4. Negative Semidefinite	$\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0, \quad \forall \mathbf{x} \neq 0$	
5. Indefinite	$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for some $\mathbf{x} \neq 0$ and $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$ for other $\mathbf{x} \neq 0$	Neither

- Examples:

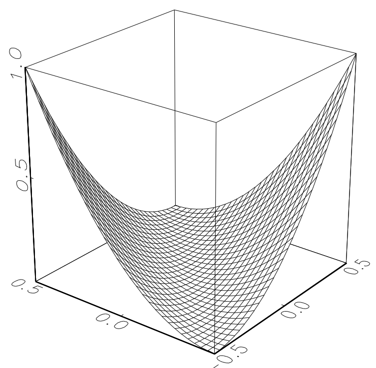
1. Positive Definite:

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x} \\ &= x_1^2 + x_2^2 \end{aligned}$$



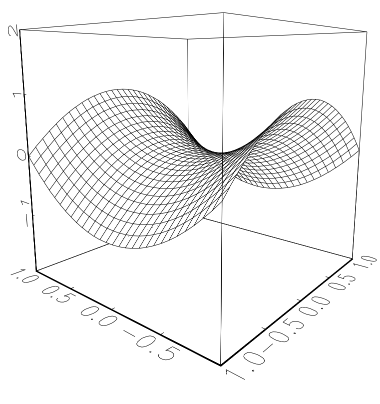
2. Positive Semidefinite:

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \mathbf{x} \\ &= (x_1 - x_2)^2 \end{aligned}$$



3. Indefinite:

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{x} \\ &= x_1^2 - x_2^2 \end{aligned}$$



3 Test for Definiteness using Principal Minors

- Given an $n \times n$ matrix $\mathbf{A}$, kth order **principal minors** are the determinants of the $k \times k$ submatrices along the diagonal obtained by deleting $n-k$ columns and the same $n-k$ rows from $\mathbf{A}$.
- Example: For a 3×3 matrix $\mathbf{A}$,
  1. First order principle minors:
  $$|a_{11}|, \quad |a_{22}|, \quad |a_{33}|$$
  2. Second order principle minors:
  $$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$
  3. Third order principle minor: $|\mathbf{A}|$
- Define the kth **leading principal minor** M_k as the determinant of the $k \times k$ submatrix obtained by deleting the last $n-k$ rows and columns from $\mathbf{A}$.
- Example: For a 3×3 matrix $\mathbf{A}$, the three leading principal minors are
$$M_1 = |a_{11}|, \quad M_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad M_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$
- Algorithm: If $\mathbf{A}$ is an $n \times n$ symmetric matrix, then
  1. $M_k > 0$, $k = 1, \ldots, n$ $\implies$ Positive Definite
  2. $M_k < 0$, for odd k and $M_k > 0$, for even k $\implies$ Negative Definite
  3. $M_k \neq 0$, $k = 1, \ldots, n$, but does not fit the pattern of 1 or 2. $\implies$ Indefinite.
- If some leading principle minor is zero, but all others fit the pattern of the preceding conditions 1 or 2, then
  1. Every principal minor ≥ 0 $\implies$ Positive Semidefinite
  2. Every principal minor of odd order ≤ 0 and every principal minor of even order ≥ 0 $\implies$ Negative Semidefinite

4 Maxima and Minima in $\mathbf{R}^n$

- **Conditions for Extrema**: The conditions for extrema are similar to those for functions on $\mathbf{R}^1$. Let $f(\mathbf{x})$ be a function of n variables. Let $B(\mathbf{x}, \epsilon)$ be the ϵ-ball about the point $\mathbf{x}$. Then
  1. $f(\mathbf{x}^*) > f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$ $\implies$ Strict Local Max
  2. $f(\mathbf{x}^*) \geq f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$ $\implies$ Local Max
  3. $f(\mathbf{x}^*) < f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$ $\implies$ Strict Local Min
  4. $f(\mathbf{x}^*) \leq f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$ $\implies$ Local Min

5 First Order Conditions

- When we examined functions of one variable x, we found critical points by taking the first derivative, setting it to zero, and solving for x. For functions of n variables, the critical points are found in much the same way, except now we set the partial derivatives equal to zero.†
- Given a function $f(\mathbf{x})$ in n variables, the **gradient** $\nabla f(\mathbf{x})$ is a column vector, where the ith element is the partial derivative of $f(\mathbf{x})$ with respect to x_i:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

- $\mathbf{x}^*$ is a critical point iff $\nabla f(\mathbf{x}^*) = 0$.
- Example: Find the critical points of $f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + 1$
  1. The partial derivatives of $f(\mathbf{x})$ are

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial x_1} &= 2(x_1 - 1) \\ \frac{\partial f(\mathbf{x})}{\partial x_2} &= 2x_2 \end{aligned}$$

  2. Setting each partial equal to zero and solving for x_1 and x_2, we find that there's a critical point at $\mathbf{x}^* = (1, 0)$.

6 Second Order Conditions

- When we found a critical point for a function of one variable, we used the second derivative as an indicator of the curvature at the point in order to determine whether the point was a min, max, or saddle. For functions of n variables, we use second order partial derivatives as an indicator of curvature.
- Given a function $f(\mathbf{x})$ of n variables, the **Hessian** $\mathbf{H}(\mathbf{x})$ is an $n \times n$ matrix, where the (i, j)th element is the second order partial derivative of $f(\mathbf{x})$ with respect to x_i and x_j:

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

- **Curvature and The Taylor Polynomial as a Quadratic Form**: The Hessian is used in a Taylor polynomial approximation to $f(\mathbf{x})$ and provides information about the curvature of $f(\mathbf{x})$ at $\mathbf{x}$ — e.g., which tells us whether a critical point $\mathbf{x}^*$ is a min, max, or saddle point.

†We will only consider critical points on the interior of a function's domain.

1. The second order Taylor polynomial about the critical point $\mathbf{x}^*$ is

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T\mathbf{H}(\mathbf{x}^*)\mathbf{h} + R(\mathbf{h})$$

2. Since we're looking at a critical point, $\nabla f(\mathbf{x}^*) = 0$; and for small $\mathbf{h}$, $R(\mathbf{h})$ is negligible. Rearranging, we get

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^T\mathbf{H}(\mathbf{x}^*)\mathbf{h}$$

3. The RHS is a quadratic form and we can determine the definiteness of $\mathbf{H}(\mathbf{x}^*)$.

   (a) If $\mathbf{H}(\mathbf{x}^*)$ is positive definite, then the RHS is positive for all small $\mathbf{h}$:

   $$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) > 0 \quad\Longrightarrow\quad f(\mathbf{x}^* + \mathbf{h}) > f(\mathbf{x}^*)$$

   i.e., $f(\mathbf{x}^*) < f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$, so $\mathbf{x}^*$ is a strict local min.

   (b) Conversely, if $\mathbf{H}(\mathbf{x}^*)$ is negative definite, then the RHS is negative for all small $\mathbf{h}$:

   $$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) < 0 \quad\Longrightarrow\quad f(\mathbf{x}^* + \mathbf{h}) < f(\mathbf{x}^*)$$

   i.e., $f(\mathbf{x}^*) > f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$, so $\mathbf{x}^*$ is a strict local max.

- **Summary of Second Order Conditions**:

  Given a function $f(\mathbf{x})$ and a point $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = 0$,

1. $\mathbf{H}(\mathbf{x}^*)$ Positive Definite	$\Longrightarrow$	Strict Local Min
2. $\mathbf{H}(\mathbf{x})$ Positive Semidefinite $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$	$\Longrightarrow$	Local Min
3. $\mathbf{H}(\mathbf{x}^*)$ Negative Definite	$\Longrightarrow$	Strict Local Max
4. $\mathbf{H}(\mathbf{x})$ Negative Semidefinite $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$	$\Longrightarrow$	Local Max
5. $\mathbf{H}(\mathbf{x}^*)$ Indefinite	$\Longrightarrow$	Saddle Point

- Example: We found that the only critical point of $f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + 1$ is at $\mathbf{x}^* = (1, 0)$. Is it a min, max, or saddle point?

  1. Recall that the gradient of $f(\mathbf{x})$ is

  $$\nabla f(\mathbf{x}) = \begin{pmatrix} 2(x_1 - 1) \\ 2x_2 \end{pmatrix}$$

     Then the Hessian is

  $$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

  2. To check the definiteness of $\mathbf{H}(\mathbf{x}^*)$, we could use either of two methods:

     (a) Determine whether $\mathbf{x}^{\mathbf{T}}\mathbf{H}(\mathbf{x}^*)\mathbf{x}$ is greater or less than zero for all $\mathbf{x} \neq \mathbf{0}$:

     $$\mathbf{x}^{\mathbf{T}}\mathbf{H}(\mathbf{x}^*)\mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1^2 + 2x_2^2$$

     For any $\mathbf{x} \neq \mathbf{0}$, $2(x_1^2 + x_2^2) > 0$, so the Hessian is positive definite and $\mathbf{x}^*$ is a strict local minimum.

     (b) Using the method of leading principal minors, we see that $M_1 = 2$ and $M_2 = 4$. Since both are positive, the Hessian is positive definite and $\mathbf{x}^*$ is a strict local minimum.

7 Global Maxima and Minima

- To determine whether a critical point is a global min or max, we can check the concavity of the function over its entire domain. Here again we use the definiteness of the Hessian to determine whether a function is globally concave or convex:

  1. $\mathbf{H}(\mathbf{x})$ Positive Semidefinite $\forall \mathbf{x}$ $\Longrightarrow$ Globally Convex
  2. $\mathbf{H}(\mathbf{x})$ Negative Semidefinite $\forall \mathbf{x}$ $\Longrightarrow$ Globally Concave

  Notice that the definiteness conditions must be satisfied over the entire domain.

- Given a function $f(\mathbf{x})$ and a point $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = 0$,

  1. $f(\mathbf{x})$ Globally Convex $\Longrightarrow$ Global Min
  2. $f(\mathbf{x})$ Globally Concave $\Longrightarrow$ Global Max

- Note that showing that $\mathbf{H}(\mathbf{x}^*)$ is negative semidefinite is not enough to guarantee $\mathbf{x}^*$ is a local max. However, showing that $\mathbf{H}(\mathbf{x})$ is negative semidefinite for all $\mathbf{x}$ guarantees that x^* is a global max. (The same goes for positive semidefinite and minima.)

- Example: Take $f_1(x) = x^4$ and $f_2(x) = -x^4$. Both have $x = 0$ as a critical point. Unfortunately, $f_1''(0) = 0$ and $f_2''(0) = 0$, so we can't tell whether $x = 0$ is a min or max for either. However, $f_1''(x) = 12x^2$ and $f_2''(x) = -12x^2$. For all x, $f_1''(x) \geq 0$ and $f_2''(x) \leq 0$ — i.e., $f_1(x)$ is globally convex and $f_2(x)$ is globally concave. So $x = 0$ is a global min of $f_1(x)$ and a global max of $f_2(x)$.

8 One More Example

- Given $f(\mathbf{x}) = x_1^3 - x_2^3 + 9x_1x_2$, find any maxima or minima.

  1. First order conditions. Set the gradient equal to zero and solve for x_1 and x_2.

  $$\begin{aligned} \frac{\partial f}{\partial x_1} &= 3x_1^2 + 9x_2 = 0 \\ \frac{\partial f}{\partial x_2} &= -3x_2^2 + 9x_1 = 0 \end{aligned}$$

  We have two equations in two unknowns. Solving for x_1 and x_2, we get two critical points: $\mathbf{x}_1^* = (0, 0)$ and $\mathbf{x}_1^* = (3, -3)$.

  2. Second order conditions. Determine whether the Hessian is positive or negative definite. The Hessian is

  $$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 6x_1 & 9 \\ 9 & -6x_2 \end{pmatrix}$$

  Evaluated at $\mathbf{x}_1^*$,

  $$\mathbf{H}(\mathbf{x_1^*}) = \begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix}$$

  The two leading principal minors are $M_1 = 0$ and $M_2 = -81$, so $\mathbf{H}(\mathbf{x_1^*})$ is indefinite and $\mathbf{x}_1^* = (0, 0)$ is a saddle point.

Evaluated at $\mathbf{x}_2^*$,

$$\mathbf{H}(\mathbf{x_2^*}) = \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}$$

The two leading principal minors are $M_1 = 18$ and $M_2 = 243$. Since both are positive, $\mathbf{H}(\mathbf{x_2^*})$ is positive definite and $\mathbf{x}_2^* = (3, -3)$ is a strict local min.

3. Global concavity/convexity. In evaluating the Hessians for $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ we saw that the Hessian is not everywhere positive semidefinite. Hence, we can't infer that $\mathbf{x}_2^* = (3, -3)$ is a global minimum. In fact, if we set $x_1 = 0$, the $f(\mathbf{x}) = -x_2^3$, which will go to $-\infty$ as $x_2 \to \infty$.